Measuring cross correlation function line-profile variations in radial-velocity measurements via a Skew Normal distribution

U. Simola X. Dumusque Jessi Cisewski-Kehe

May 8, 2018

Abstract

When working with radial velocities for detecting Extrasolar planets and when using data from stabilized spectrographs, the different moments of the cross-correlation function (CCF) are used to measure the radial velocity of the star and also changes in the shape of the CCF. Those changes are due to stellar activity and therefore the best precision on these moments is required to de-correlate exoplanet signals from spurious RV signals originating from stellar activity.

We propose here to measure those moments using a Skew Normal (SN) distribution, which compared to the Normal distribution generally used, naturally provides an extra parameter to model the CCF natural asymmetry. Moreover, since the CCF presents a natural asymmetry due the convective blueshift, the SN distribution seems to provide a more realistic model to the CCF.

We analyze 5 stars with different activity levels and different signal-to-noise ratio levels. In each case, we compare the results obtained by fitting to the CCF respectively a Normal and a SN. We also estimate rigorous errors for the different moments of the CCF using a bootstrap analysis.

The correlations between the RVs and the CCF asymmetry or RVs and the CCF width is always stronger when using the parameters derived from the SN. Therefore all the moments of the CCF are more sensitive to activity, which is a huge gain to probe stellar activity. RVs derived from a SN distribution are more sensitive to activity, which is interesting when looking for rotational periods or characterizing better stellar activity. [[**Referee:** " interesting" is a highly subjective and not very precise or quantitative expression. Interesting in what sense? I would term this a "sentence of no value", i.e. one that conveys not useful information to the reader.]] The precision on the asymmetry parameter derived from the SN is on average 15% better than the one calculated on the common Bisector Inverse Slope Span (BIS SPAN). When using the median of the SN as parameter defining the radial velocity of the star, the errors on the RVs measured by the SN are on average 10% smaller than the errors calculated on the mean of the Normal.

We strongly encourage the use of the SN distribution to retrieve in 1 single fit the different moments of the CCF, because the derived moments probe better stellar activity signals than when using a Normal distribution, leading as well to smaller standard errors associated to the radial velocity and the asymmetry parameters. [[**Referee:** You simply cannot make this claim. What you have shown is that the SN method produces errors that are 60% larger than the Normal distribution

and you hint that this is due to it being more sensitive to activity. You have not demonstrated this by looking at the ”activity” signal that the SN isolates to see if it is actually more sensitive at finding rotation periods, activity cycles, etc. Where is the sensitivity in actually getting out useful information? Furthermore, after correcting the SN RVs for the activity, you arrive at the same answer as the Normal distribution method. I would conclude that the SN offers no real advantages over Normal so the community can continue to use the standard method. Grammatically this is an ambiguous statement. Are you looking for ”better signals”, i.e. ones that you like, or probing stellar signals in a better way? (This is what you meant, but not what was stated.)]]

1 Introduction

[[**Jessi:** I added subsections here for our reference to make sure the bases are covered, but we should remove them before submission.]]

1.1 Goal of RV analysis

The radial velocity (RV) of a star is defined to be the velocity of the center of mass of the star along our line of sight. [[**Referee:** No, the radial velocity you measure of the star has many components that contribute to the RV: the space motion (present even if there is no companion), the motion about the center of mass (the so-called ?-velocity), oscillations, convective blue shift, etc. You are just assuming that you are measuring the reflex motion of a star-planet system. There are many physical processes that produce a Doppler shift besides exoplanets.]] This quantity can be derived precisely by measuring the Doppler shift of spectral lines produced in stellar atmospheres. [[**Referee:** Well, not if you have a poor measurement precision. You only get a ”precise” measurement with simultaneous wavelength calibration. Again, the authors are not being very ”precise” here.]] For spectrographs that are not stabilised in temperature and pressure, the iodine technique is used, where the light of the star passes through a iodine cell before getting into the spectrograph to imprint the absorption spectrum of iodine on top of the stellar spectrum (The Hamilton spectrograph (Vogt 1987) at Lick Observatory, HIRES (Vogt et al. 1994) on the Keck 10-m telescope, the Tull spectrograph (Tull et al. 1995), the High Resolution Spectrograph HRS (Tull 1998)). [[**Jessi:** we should mention the new spectrographs like EXPRES]] In this case, if the spectrograph shifts due to changing atmospheric conditions, the iodine and stellar spectra are shifted in the same way. This leads to some complications when reducing the data because one has to decorrelate the iodine spectrum from the stellar spectrum. [[**Referee:** This is a poorly expressed thought that is technically wrong. The spectrograph does not shift due to changing atmospheric conditions (earth’s atmosphere?). It changes because of mechanical shifts (vibrations, temperature changes in spectrograph housing, changes in instrumental profile) as well as movement of the image on the slit or fiber which yes, does affect stellar and iodine lines in the same way.]] [[**Referee:** No, you do not decorrelate (and I am not sure what the authors mean by this word) the iodine from the stellar spectrum. The authors show that they have no understanding how the method works. What is actually done is that a high-resolution iodine spectrum is combined with a spectrum of the star without iodine lines and a fit is made to the observed star+iodine spectrum. It is not decorrelation, but rather χ^2 fitting.]]

1.2 Stellar activity effect on CCF, Normal fit plus FWHM indicator

[[**Jessi:** Eventually three methods are introduced for detecting CCF asymmetry: FWHM, bisector span, and the Figueira approach - it seems that these methods need to be better explained. Somewhere in the intro it needs to become clear that we are (i) providing a new way of measuring CCF asymmetry, (ii) this is a unified approach that also provides an RV estimate, and (iii) it (hopefully) provides a better indicator.]] For spectrographs that are stabilized, the spectrum of a calibration lamp is recorded close to the stellar spectrum on the CCD, which prevents contamination of the stellar spectrum (CORALIE (Queloz et al. 2000), The High Accuracy Radial Velocity Planet Searcher (HARPS) (PHASE 2003), HARPS- N (Cosentino et al. 2012), SOPHIE (Bouchy et al. 2013), CARMENES (Quirrenbach et al. 2014)). For those instruments, reducing the data is easier as the stellar spectrum is not contaminated with iodine absorption lines. [[**Referee:** This is not true! I have taken spectra with HARPS and there is cross-talk between spectra and Th-Ar particularly for strong emission lines. So there is some contamination. Furthermore, this "preventing of contamination" as the authors state has nothing to do with the spectrograph being stablized, it has to do with the simultaneous wavelength calibration using two fibers. In fact one reference the authors give, CORALIE, is not a stabilized spectrograph!]] [[**Referee:** In paragraph 1 you give two lists of spectrographs. I presume the first one is a list spectrographs that use the iodine method, although that is not explicitly expressed. The authors talk about the iodine cell and then just give a parenthetical list of spectrographs. There is no explicit statement that these spectrographs, designed for other purposes, are all equipped with an iodine cell. The reader must assume this. Furthermore, the list is not all-inclusive. I can think of a dozen facilities that use iodine cells. So you should say, "For example" or..."to name a few" The second list it is for simultaneous Th-Ar calibration. Not all of these are "stabilized" like the authors claim. CORALIE, if I am not mistaken, was simultaneous Th-Ar without stabilization. That is why they built HARPS. Furthermore, all of theses spectrographs were designed for precise stellar RVs, unlike the spectrographs listed for the iodine method.]]

For stabilized spectrographs, the RV is derived by first correlating a stellar spectrum at a particular time with a synthetic (Baranne et al. 1996; Pepe et al. 2002) or an observed stellar template (Anglada-Escudé and Butler 2012), which gives an average line profile, generally called Cross Correlation Function (CCF). The CCF is a function of RV, and the RV corresponding to the minimum of the CCF is the estimated RV for that spectrum. However, this would give a noisy estimate so first a Normal density shape is fitted to the CCF, and then the estimated mean of the Normal density is used as the estimate of the RV. The full width at half maximum (FWHM) of the Normal fit is also used as an indicator of spectral line asymmetry. The CCF technique allows for an averaging of thousands of spectral lines and can therefore reach a very high signal-to-noise ratio (SNR), which is essential for a good RV precision.

The convection in external layers of solar type stars is responsible for the granulation pattern than can be seen at high spatial resolution on the surface of the Sun. The differences in flux and velocity between upflows and downflows change the Normal profile of spectral lines that become asymmetric with a "C"-shaped profile (Dravins et al. 1981). [[**Referee:** Regarding the C-shape of the bisector. The classic C-shape is only for late-type stars. In fact for hotter stars (F-type) it reverses. The strength of the asymmetry depends not just on the velocities, it is more complicated than that. It also depends on the flux ratio between hot cells and cool lanes, as well as the ratio of surface areas between the two.]] The strength of the asymmetry depends on the velocity of the convection, approximatively $300\,\mathrm{m\,s^{-1}}$for the Sun, but also on the formation depth of spectral lines (Gray 2009). Since the CCF is an average of all the spectral lines, some strongly asymmetric and some not, its asymmetry is rather small, which is why a Normal distribution can be an appropriate model for the CCF. This small asymmetry modifies however slightly the estimated RVs of the star, reducing the accuracy of the measurement, but if this asymmetry does not vary with time, the precision is kept. [[**Referee:** But the activity causes the asymmetry to change due to rotation, spot evolution, etc. The authors are confusing the readers here. They talk about a changing activity signal, but now argue that the precision is only kept if the asymmetry is constant, which it is not. I think what they want to say is that you measure a precise stellar radial velocity, but the activity makes its own contribution which reduces the accuracy of the RV determination for the barycentric motion due to a planet.]]

Convection is not the only phenomenon responsible for asymmetries in the single spectral line and the CCF. Stellar activity is responsible for the appearance of dark spots and bright faculae on the stellar photosphere, which breaks the flux balance between the red-shifted and the blue-shifted halves of a rotating star and therefore induce an asymmetry of spectral lines and thus of the CCF. As the star rotates, spots and faculae move across the stellar disk, modifying the line asymmetry and thus producing an apparent Doppler shift (Saar and Donahue 1997; Hatzes 2002; Kurster et al. 2003; Desort et al. 2007; Lagrange et al. 2010; Boisse et al. 2012). Spots and faculae are also regions where the magnetic field is strong. Strong magnetic fields reduce stellar convection, which in turn modifies the asymmetry of spectral lines (Cavallini et al. 1985; Dravins et al. 1981; Lindegren and Dravins 2003; Meunier et al. 2010; Dumusque et al. 2014).

1.3 Bisector span indicators of stellar activity

[[**Jessi:** Its seems that more details should be added here]] Stellar activity can induce RV variations by a modification of the spectral line asymmetry, while any orbiting companions would induce a pure Doppler shift on *all* spectral lines without modifying their shape. Therefore, assuming that there is no instrumental systematics, stellar activity generally induces a variation in line asymmetry that can be observed in the FWHM of the CCF. The line asymmetry is commonly retrieved by calculating the bisector of the CCF (Voigt 1956) and deriving the bisector curvature (Hatzes 1996) or the bisector inverse slope span (BIS SPAN, Queloz et al. 2001). An orbiting exoplanet will produce an RV variation induced by a pure Doppler shift of *all* the spectral lines of each stellar spectrum. Stellar activity, however, does not produce a blueshift or redshift of the spectra, but can create a spurious RV signal by modifying the shape of spectral lines. To track changes in line shape, the FWHM, the BIS SPAN or the indicators introduced by Figueira et al. (2013) are often used, which provide information on the average width and asymmetry of the CCF (Hatzes 1996; Fiorenzano et al. 2005; Queloz et al. 2001). A strong correlation between the estimated RVs and one or more of these parameters is a sign that the estimated RVs may be contaminated by stellar activity. [[**Referee:** It is the slope or span, not both. And you generally do not calculate the inverse, just the span. This may be correlated or inversely correlated with the RV.]]

1.4 Figueira indicators of stellar activity

Figueira et al. (2013) proposed different indicators, including a bi-Gaussian fitting of the CCF [[**Jessi:** this needs to be better explained]].

1.5 Issues with detecting stellar activity

Unfortunately, when analyzing slow rotators stars such as the Sun, due to the limited spectral resolution of the spectrographs and the limited precision in RV, it becomes difficult to measure the line asymmetry, resulting in complications for detecting very small-mass planets with the RV technique. In the procedure described above, the measurement of the RV and the FWHM is done separately from the measurement of the line asymmetry. All those parameters are correlated when stellar activity is dominant, and performing a step-by-step approach makes it difficult to correctly derive the errors on the different parameters retrieved. [[**Referee:** I simply do not understand this statement, possibly because it does not follow a standard English construction. You have a CCF and you measure a width and bisector span. This is a measurement error determined by the S/N, the resolution, and the stability of the spectrograph. You can determine an error in a quantity (RV and BIS). You can also have an uncertainty in the contribution of activity to the RV signal, but that is not a formal error. True errors come from photon statistics, systematic errors, instrumental errors, etc. You can attempt to remove the signal due to activity and that has an associated error, but that error is due to your lack of knowledge of the surface structure causing this RV and the exact value of the contribution to the RV. This is different from a measurement error. Maybe that is what you mean here, but it is not clear from what is written.]] In addition, the Normal distribution cannot take into account the natural asymmetry of the CCF, leaving correlated noise in the residuals, which also complicates the determination of errors. We propose to overcome these problems by fitting a SN (SN) distribution to the CCF, which naturally includes a skewness parameter (Azzalini 1985).

1.6 Outline of paper

[[**TODO:** This paragraph will need to be updated once the rest is finished.]] The paper is organized as follow. In Sec. 2 we introduce the SN distribution and describe its applicability for modeling the CCF, show that the SN density is a better representation of observed CCF than a Normal distribution, and study how the SN parameters relate to the RV, FWHM and BIS SPAN of the CCF. In Sec. 1 we present a simple model to correct for stellar activity. In Sec. 4, we compare on real observations the sensitivity of the SN parameters to stellar activity with respect to other existing indicators. In Sec. 5 we derive error bars for the different CCF parameters, and finally we discuss our results and conclude in Sec. 6 and Sec. 7.

2 The Skew Normal distribution

The Skew Normal (SN) distribution is a class of probability distributions which includes the Normal distribution as a special case (Azzalini 1985). The SN distribution has, in addition to a location and a scale parameter analogous to the Normal distribution's mean and standard deviation, a third parameter which describes the asymmetry, or the skewness, of the distribution. Considering a random variable $Y \in \mathbb{R}$ (where $\mathbb{R}$ is the real line) which follows a SN distribution with location parameter $\xi \in \mathbb{R}$, scale parameter $\omega \in \mathbb{R}^+$ (i.e., the positive real line), and skewness parameter $\alpha \in \mathbb{R}$, its density at some value $Y = y$ can be written as

$$SN(y; \xi, \omega, \alpha) = \frac{2}{\omega}\phi\left(\frac{y-\xi}{\omega}\right)\Phi\left(\frac{\alpha(y-\xi)}{\omega}\right) \tag{1}$$

where ϕ and Φ are respectively the density function and the distribution function of a *standard* Normal distribution[1] and $\alpha \in \mathbb{R}$ is the skewness parameter which quantifies the asymmetry of the SN. We then write $Y \sim SN(\xi, \omega^2, \alpha)$ to mean that the random variable Y follows the noted SN distribution. Examples of SN densities under different skewness parameter values and the same location and scale parameters ($\xi = 0$ and $\omega = 1$) are displayed in Fig. 1. A usual Normal distribution is the special case of the SN distribution when the skewness parameter, α, is equal to 0.[2] For reasons related to the interpretation of the parameters in Equation (1) and computational issues with estimating α near 0, a different parametrization is used, which is referred to as the *centered parametrization* (CP). We will be using the CP in this work, which includes a mean parameter μ, a variance parameter σ^2, and a skewness parameter γ. In order to define the CP, we need to express the CP parameters (μ, σ^2, γ) as a function of the one used in the Equation (1) with (ξ, ω^2, α) by

$$\mu = \xi + \omega\beta, \quad \sigma^2 = \omega^2(1-\beta^2), \quad \gamma = \frac{1}{2}(4-\pi)\beta^3\left(1-\beta^2\right)^{-3/2} \tag{2}$$

where $\beta = \sqrt{\frac{2}{\pi}}\left(\frac{\alpha}{\sqrt{1+\alpha^2}}\right)$.

By using Eq. 2, the new set of parameters (μ, σ^2, γ) provides a more clear interpretation of the behavior of the SN distribution. For the α values used in Fig. 1, the corresponding values of μ, σ^2, γ are displayed in Table 1. In particular, μ and σ^2 are the actual mean and variance of the distribution (rather than simply a location and scale parameter), and γ becomes an index for evaluating the asymmetry of the SN.

[1] A standard Normal distribution is a Normal distribution with a mean of 0 and a standard deviation of 1.

[2] This can be seen from Equation (1). If $\alpha = 0$ then $\Phi\left(\frac{\alpha(y-\xi)}{\omega}\right) = \Phi(0)$; this is the the probability a standard Normal random variable is less than or equal to 0, which is 0.5. The 0.5 cancels with the 2 in the denominator and what remains is the usual Normal density, $\frac{1}{\omega}\phi\left(\frac{y-\xi}{\omega}\right)$

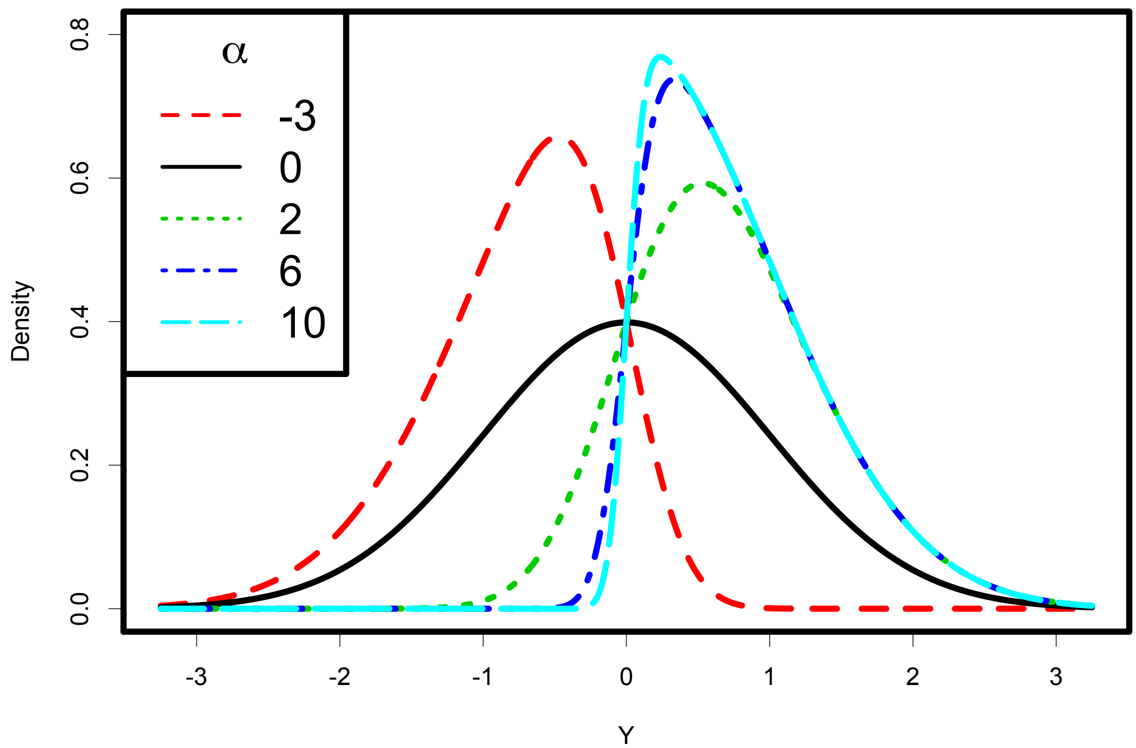

Figure 1: Density function of a random variable $Y \sim SN(\xi, \omega^2, \alpha)$ with location parameter $\xi = 0$, scale parameter $\omega = 1$, and different values of the skewness parameter α indicated by different colors and line types. Note that the solid black line has an $\alpha = 0$ making it a Normal distribution.

Beyond the mean of the SN, it is convenient to introduce a second location parameter that will be largely used in the analyses: the median. The median of the SN is defined as that value m such that

$$\int_{-\infty}^{m} SN(y;\xi,\omega,\alpha) = \frac{1}{2}, \tag{3}$$

where $SN(y;\xi,\omega,\alpha)$ follows Equation 1.[3]

α	μ	σ^2	γ
-3	-0.757	0.427	-0.667
0	0.000	1.000	0.000
2	0.714	0.491	0.454
6	0.787	0.381	0.891
10	0.794	0.370	0.956

Table 1: CP values, (μ,σ^2,γ), corresponding to the α values from Fig. 1 (with location parameter $\xi = 0$ and scale parameter $\omega = 1$) using Equation (2). Values are rounded to three decimal places.

Further details about the parametrization from Equation (1) (called *Direct Parametrization* or DP), the CP, and general statistical properties of the SN are treated in rigorous mathematical and statistical viewpoints in the book by Azzalini and Capitanio (2014).

2.1 Fitting the Skew Normal distribution to the CCF

The SN density shape is used to model the CCF. In particular, the following function is fit using least-squares estimation,

$$f_{CCF}(y_i) = \mathrm{C} - \mathrm{A} \times SN(y_i;\mu,\sigma^2,\gamma), \quad i = 1,\ldots,n \tag{4}$$

where C is an unknown offset fitting the continuum of the CCF, A is an unknown amplitude parameter, and $y_1,\ldots,y_n$ are the set of RV's considered for the CCF. Note that the CCF is expressed in flux as a function of the lag of the cross-correlation template, expressed in RV.

Since the CCF presents a natural asymmetry due the convective blueshift, the SN distribution can better catch this aspect, together with other changes in asymmetry, respect the Normal fit. To check this assumption, we compared the CCF residuals after fitting a Normal and a SN distribution for 2 stars. The first star is Alpha Centauri b, whose CCFs have high signal-to-noise ratio (SNR). The second star is Corot-7, whose CCFs have low SNR. Fig. 2 shows that the SN seems to be a slightly better model to explain the shape of the CCF, in particular as the SNR decreases.

In the following of the paper, we define RV as the mean of the Normal. Concerning the fit of the CCF using the SN, we present at first 2 indicators that define the RV of the star: the mean of the SN, defined as SN mean RV and the median of the SN, defined as SN median RV (i.e.

[3]We recall that when using a symmetric distribution such as the Normal distribution, the mean and the median are equivalent.

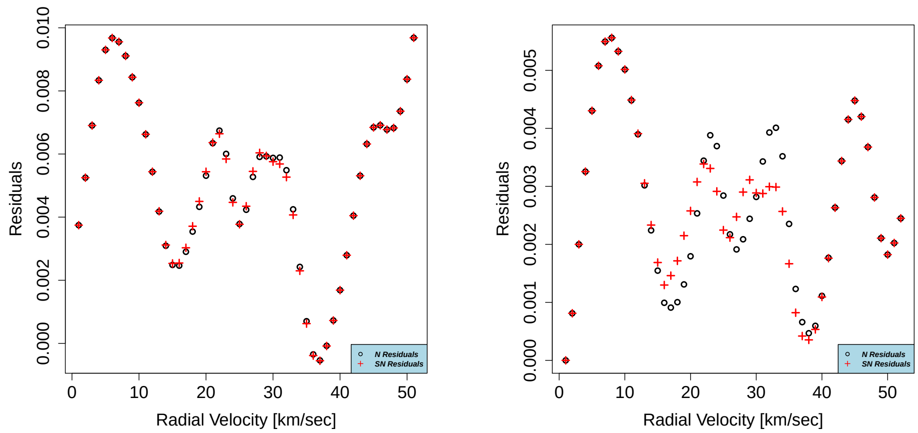

Figure 2: Comparison between the Normal (black circles) and the SN (red crosses) residuals using CCFs of the star Alpha Centauri b (left) and Corot-7 (right). The results are comparable looking at the residuals corresponding to the tails of the CCF but the SN fit leads to slightly better results in the center of the CCF for both the stars. As the signal-to-noise decreases, as happens when analyzing the star Corot-7, the SN distribution leads to better residuals respect using the Normal.

looking at Equation (3), m = SN median RV). We will discuss advantages and limits for both these choices in the examples. For the width of the CCF, we use the FWHM of the Normal, which is $2\sqrt{(2\ln 2)}\sigma$. The width of the SN, SN FWHM, is defined in the same way[4]. Being a Normal distribution symmetric, there is not such a parameter that evaluates the asymmetry of the distribution, so the BIS SPAN is used. This will be compared to the asymmetric parameter γ of the SN, called SN GAMMA. To test the strength of the correlation between the RVs and the different activity indicators, we calculated the Pearson correlation coefficient, R. A $p-$value for the statistical test having null hypothesis $H_0 : R = 0$ (i.e., no correlation) is provided, along with a 95 % confidence interval for R.

3 Simulation Study

In order to evaluate the performance of the proposed SN approach for modeling the CCF, we begin by considering a simulation study using spectra generated from the Spot Oscillation And Planet (SOAP) 2.0 code (Dumusque et al. 2014)...

3.1 Faculae

[4]Note that SN FWHM does not correspond to the width of the SN distribution at half maximum like in the Normal case.

[[**Umberto:** specify characteristic of the faculae]]

Figure 3 shows the results obtained when a faculae is present on the photosphere of the star. It is possible to note that, although there is not a planet, the presence of the faculae leads to spurious variations in RVs for all the proposed indicator. The evaluation of the standard errors, which details will be discussed in Section 5, shows that the RV defined as mean SN has the largest variability. The standard errors associates with median SN and mean Normal are comparable, although the standard errors retrieved when the RV is defined as the median of the SN are 10% smaller.

Since in this case we know that the variations in RV are only caused by the faculae, the evaluation of the correlation between the set of RVs and the asymmetry parameters is important. Figure 4 summarises the correlations obtained using as RV respectively mean SN, median SN and mean Normal. The γ parameter shows a stronger correlation with the median SN (-0.944) than what the other asymmetry indicators do with their corresponding RVs. The correlation between the CCF width and the RVs is also stronger when fitting a SN rather than a Normal.

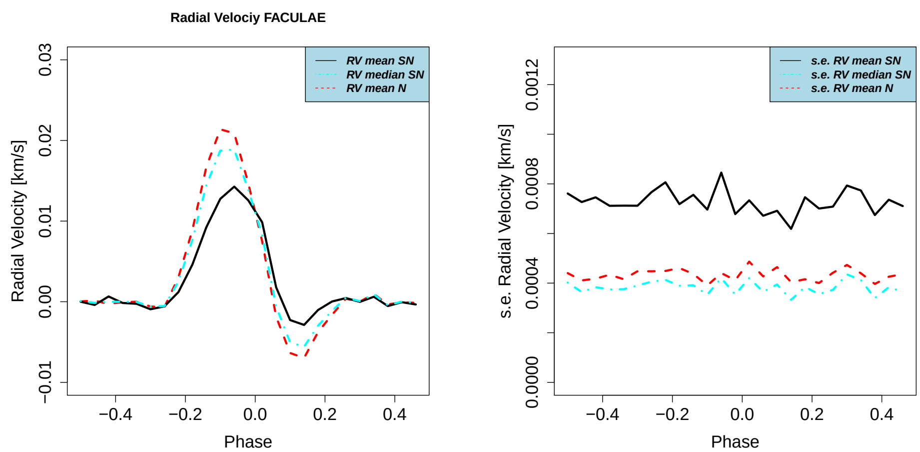

Figure 3: (left) RVs changes as function of the phase in the case in which a faculae is present on the photosphere of the star. (right) Evaluation of the standard errors corresponding to the defined RVs.

3.2 Spot

[[**Umberto:** specify characteristic of the spot]]

Figure 5 shows the results obtained when a spot is present on the photosphere of the star. Similarly to the previous case, the presence of the spot leads to spurious variations in RVs for all the proposed indicator. The evaluation of the standard errors shows that the RV defined as mean SN has the largest variability and, as before, the standard errors associated with the median SN are 10% smaller than the standard errors associated with the mean Normal.

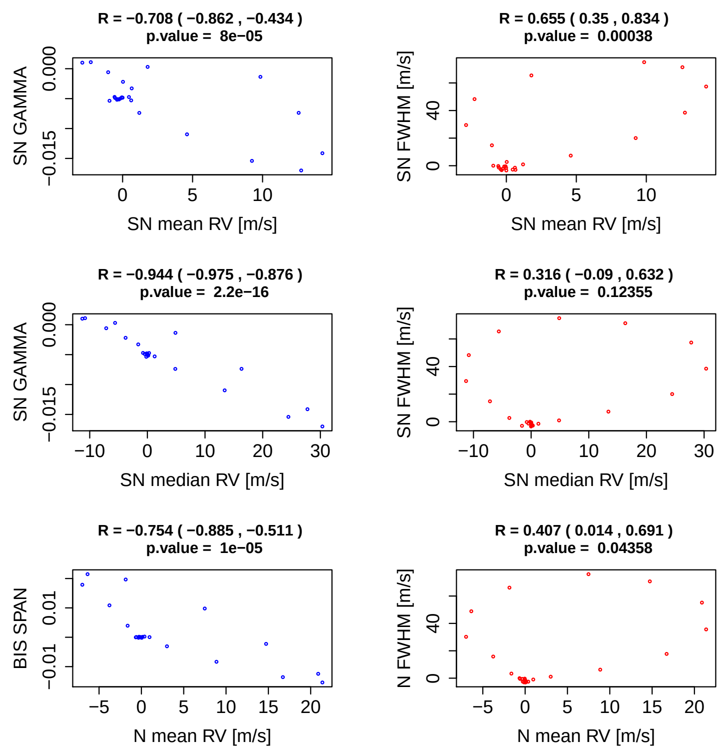

Figure 4: Evaluation of the correlation between the RVs and the asymmetry parameters when a faculae is present on the photosphere of the star.

Figure 6 summarises the correlations obtained using as RV respectively mean SN, median SN and mean Normal. It is possible to see that the correlation between the asymmetry parameters and the RVs are weaker than the previous case, because of the presence of the planet. The correlation between the CCF width and the RVs is slightly stronger when fitting a SN rather than a Normal, although in this latter case the presence of the spot is largely masquerade from the planet.

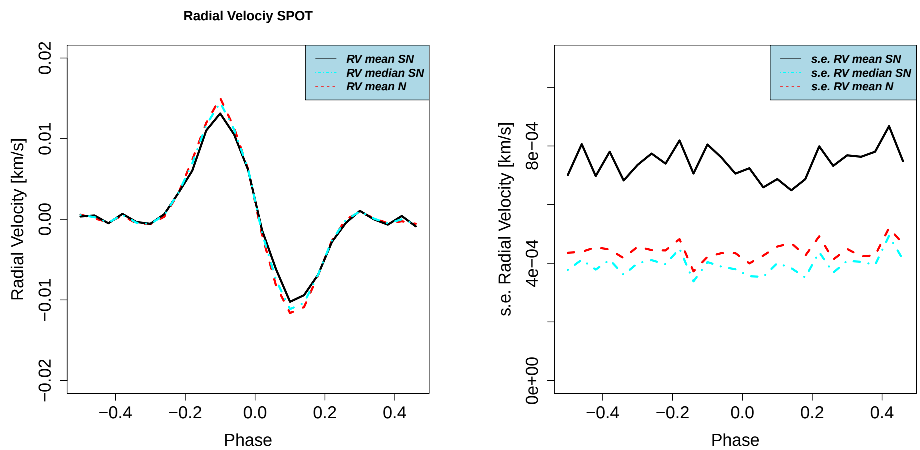

Figure 5: (left) RVs changes as function of the phase in the case in which a spot is present on the photosphere of the star. (right) Evaluation of the standard errors corresponding to the defined RVs.

3.3 Spot and planet

Figure 7 shows the results obtained when not only a spot is present on the photosphere of the star but, respect to the previous case, a planet is injected. The planet has amplitude equal to 10 $\mathrm{m\,s^{-1}}$. The evaluation of the standard errors shows that the RV defined as mean SN has the largest variability and, as before, the standard errors associated with the median SN are 10% smaller than the standard errors associated with the mean Normal.

Figure 8 summarises the correlations obtained using as RV respectively mean SN, median SN and mean Normal. The γ parameter shows a stronger correlation with the median SN (-0.89) than what the other asymmetry indicators do with their corresponding RVs. The correlation between the CCF width and the RVs is also stronger when fitting a SN rather than a Normal.

As final step, we removed to the RVs obtained in this case and shown in Figure 7(left), the signals coming from the spot, whose spurious RV effects were displayed in Figure 5. The results, presented in Figure 9, show that the periodic pure doppler shift caused by the planet is detected.

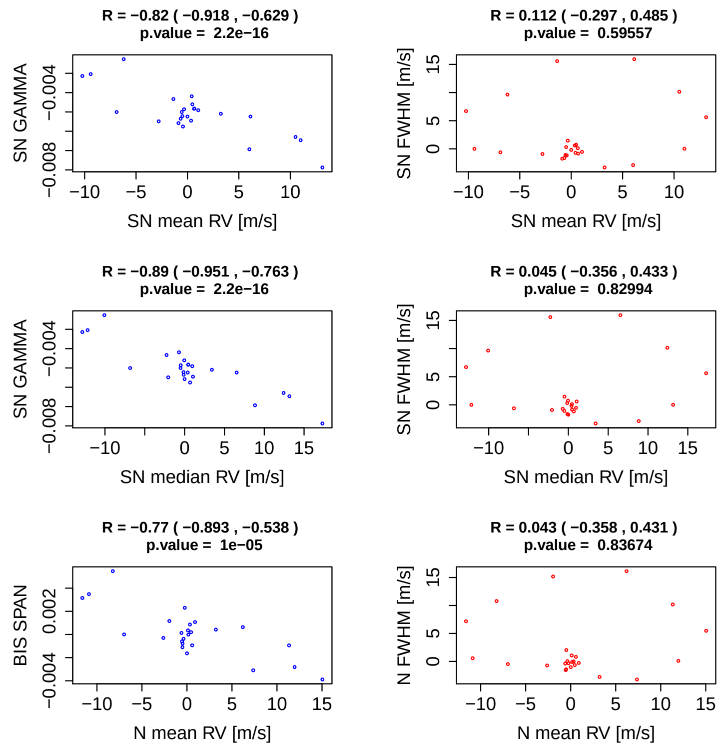

Figure 6: Evaluation of the correlation between the RVs and the asymmetry parameters when a spot is present on the photosphere of the star.

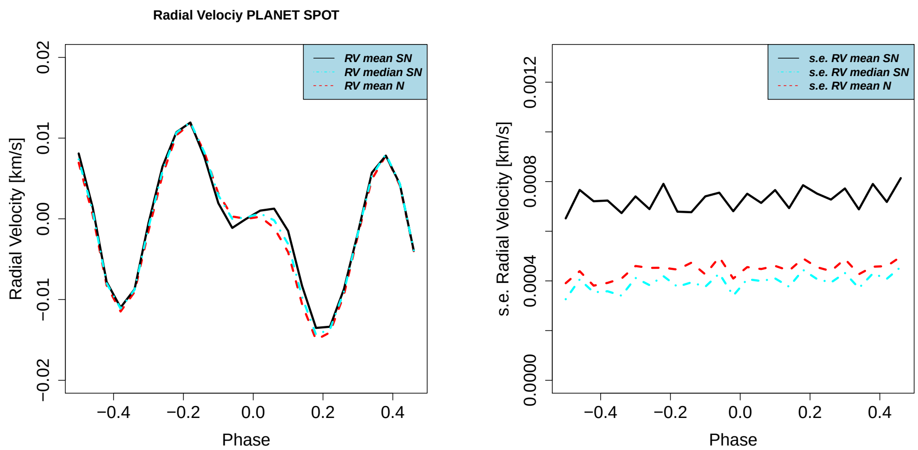

Figure 7: (left) RVs changes as function of the phase in the case in which a spot is present on the photosphere of the star and a planet is injected. (right) Evaluation of the standard errors corresponding to the defined RVs

3.4 Doppler shift added to Alpha Centauri B

We also consider a real-data example using HARPS spectra from the star Alpha Centauri B with an imputed Doppler shift added...

[[**Umberto:** to be done]]

4 Real data applications

In this Section we present the analyses conducted on Alpha Centauri b, comparing the result of fitting a CCF using the SN distribution defined in Section 2.1 with the Normal approach. Other 4 stars have been analyzed with the proposed method and details can be found in the Appendix A. For all the stars that have been considered in the present work, we selected those CCFs having SNR larger than 10.

A comparison with the results obtained by the classic approach is done, where the RVs of the star are estimated by retrieving the mean of the Normal distribution used to fit the CCF, along with the BIS SPAN or the other asymmetric parameters defined in Figueira et al. (2013). The latter parameters are calculated separately from the Normal fit that leads to the set of RVs of the star.

4.1 Alpha Centauri B

A total of 1808 CCFs measured in 2010 have been analysed from the star Alpha Centauri B. Several measurement in 2010 are contaminated by light from Alpha Centauri A. To remove contaminated

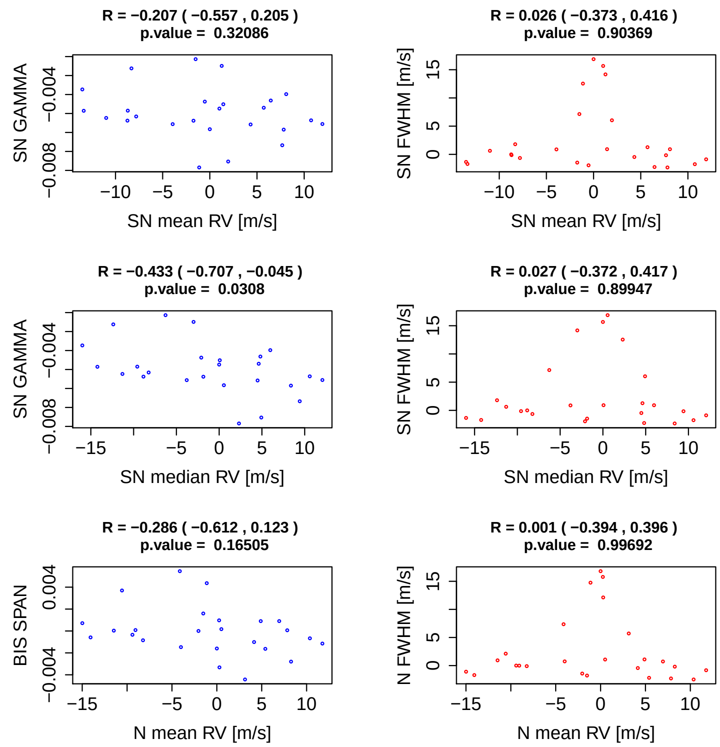

Figure 8: Evaluation of the correlation between the RVs and the asymmetry parameters when a spot is present on the photosphere of the star and a planet is injected.

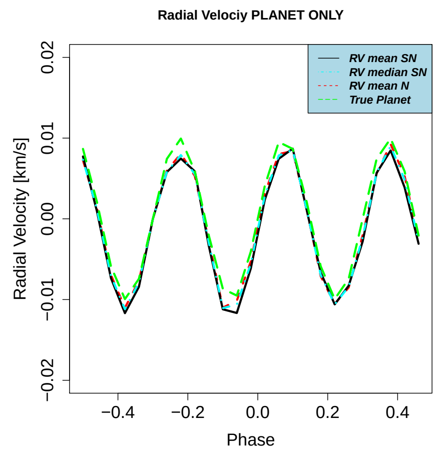

Figure 9: RVs changes as function of the phase after removed the spurious effect of the spot. In green the real signal of the planet is displayed.

spectra and thus CCFs, we performed the same selection as presented in Dumusque et al. (2012). Moreover, as noted in Dumusque et al. (2012) and Thompson et al. (2017), this dataset presents a strong stellar activity signal.

We begin the analyses by evaluating the correlation between γ and the BIS SPAN. In Fig. 10, we see that the relationship between γ and the BIS SPAN is linear, with a slope equal to 0.72 and a strong Pearson correlation coefficient of 0.954. This comparison is useful because γ is an adimensional parameter taking information about the asymmetry of the SN while the BIS SPAN, beyond this, has got unit of measure of $\mathrm{km\,s^{-1}}$. In other words, by using Fig. 10, it is possible to provide a physical meaning to γ.

Fig. 11 shows the comparison between the RVs retrieved using the SN shape and the ones obtained with the Normal shape. It is possible to appreciate the presence of a strong stellar activity signal, as expected (Dumusque et al. 2012; Thompson et al. 2017). When using SN mean RV, it is possible to observe more variations than the one present in the RVs measured by the Normal fitting. In Section 5 we will discuss in detail the reasons leading to this aspect. This happens because the mean of the SN is more sensitive to stellar activity. In fact, because the SN includes an asymmetry parameter, a RV defined as SN mean gets more shifted in the direction of the asymmetry induced by stellar activity. On the other hand, when using SN median RV, smaller variations in RV are caused by changes in the asymmetry of the CCF, because this second location parameter is a more robust indicator than the mean. The bottom plot of Fig. 11 captures this aspect. Both indicators can be used to capture and summarise the different information available in the CCF, as will be shown in the remainder of this work.

Similar to Figueira et al. (2013), we compare the correlation between the different activity indicators and the RVs of the star in Fig. 12. The correlation between γ and SN mean RV and the

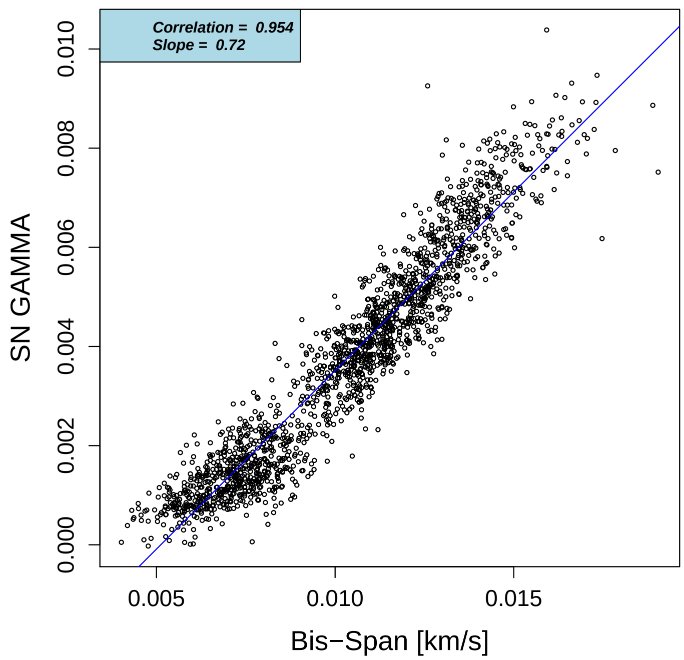

Figure 10: Correlation between γ and the BIS SPAN for Alpha Centauri B. Because γ is adimensional, retrieving the slope between γ and the BIS SPAN, which is expressed in $\mathrm{km\,s^{-1}}$, allows us to provide physical meaning to γ.

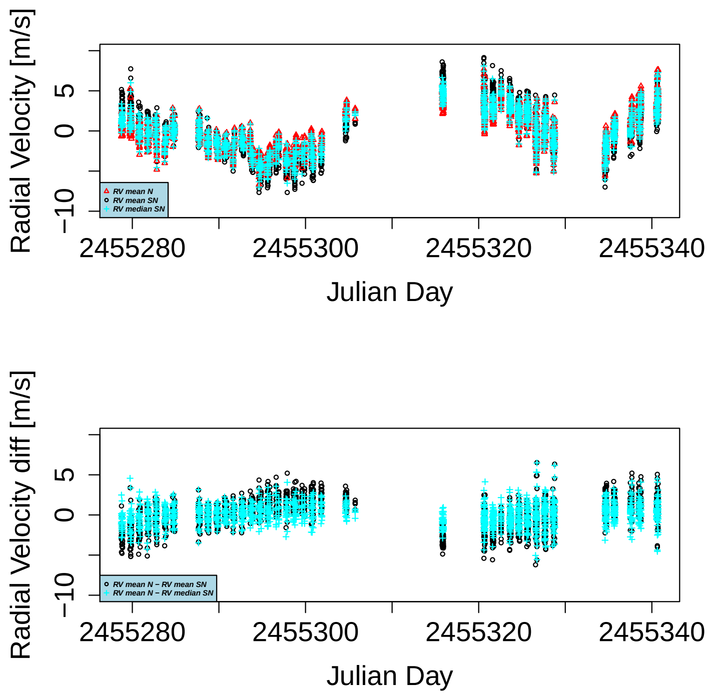

Figure 11: (top) RVs as function of Julian Day for Alpha Centauri b. The RVs are retrieved using the mean of the Normal (red triangles), SN mean RV (black circles), SN median RV (cyan crosses). (bottom) RV differences between Normal RV and SN mean RV (black circles) and between Normal RV and a SN median RV (cyan crosses).

correlation between γ and SN median RV are much stronger than the correlation calculated between the other asymmetry parameters and their corresponding RVs. In particular the correlation between γ and SN mean RV is almost twice the correlation between the other asymmetry parameters and their corresponding RVs.

Because the median is a more rubust index than the mean, the correlation between γ and SN median RV is not as large as the correlation between γ and SN mean RV, but it is nonetheless 1.5 times larger than the correlation between the other common asymmetry parameters and their corresponding RVs. In other words, changes in the asymmetry of the CCF are better captured when using the SN mean RV. The correlation between FWHM and the RVs, either by using SN mean RV or SN median RV, is as well stronger when fitting a SN distribution rather than a Normal. All the correlations are statistically different from 0.

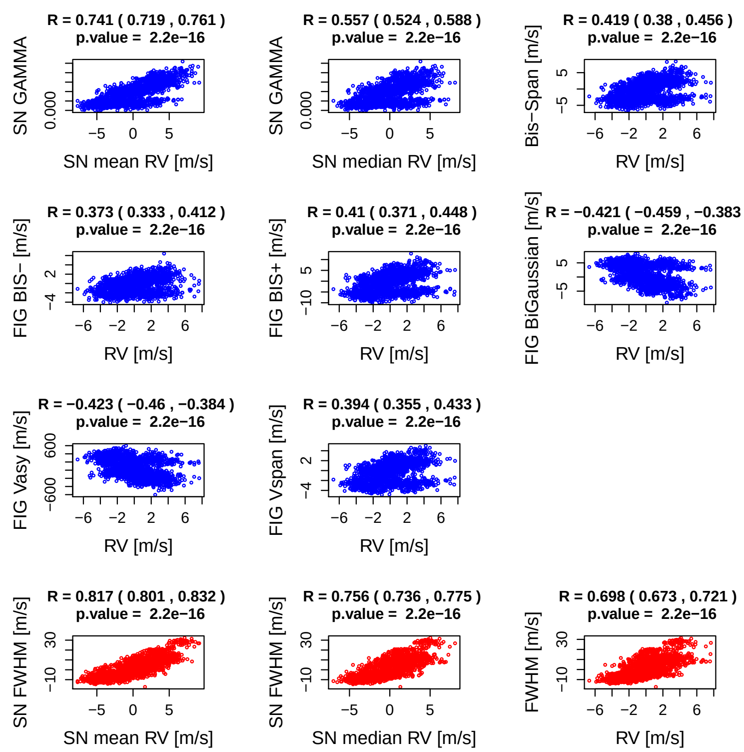

Figure 12: Correlation between the asymmetry parameters and the RVs for Alpha Centauri B. The last three plots show the correlation between the FWHM and the RVs for Alpha Centauri B using respectively the SN (SN mean RV and SN median RV) and the Normal fits.

Both the proposed indicators coming from the SN distribution have desirable and undesirable behaviors: the SN mean RV better catches changes in the asymmetry of the CCF but the resulting set of RVs ends up being contaminated by those spurious shifts caused by stellar activity that have been shortly presented in Section 1. When using SN median RV, the final set of radial velocity is less affected by those spurious shifts caused by stellar activity, but at the same time this indicator is not able to catch as well as SN mean RV changes in the shape of the CCF. Anyway, both SN mean RV and SN median RV are useful to catch different aspects of the CCF and our suggestion is to use SN mean RV when interested to retrieve information about changes in the shape of the CCF. In order to provide a set of RVs containing the smallest amount of spurious contamination imputable to stellar activity, our suggestion is instead to use SN median RV.

In the next Section we further motivate the reasons for defining as RV the median of the SN by retrieving the standard errors associated with the RV defined as the mean of the Normal, SN mean and SN median.

5 Estimation of standard errors for the CCF parameters

In this section, we perform a bootstrap analysis to measure the standard errors associate to the RVs, SN mean RVs and SN median RV, the FWHM or SN FWHM and the BIS SPAN or γ. Because a CCF is obtained from a cross-correlation, each point in a CCF is correlated with each other. Therefore, we cannot do a bootstrap analysis on perturbing independently each CCF point with a Gaussian distribution scaled to the error of each given point. To bypass this problem, we bootstrap a hundred times the stellar spectrum given the photon-noise error of each wavelength and calculate for each realization a new CCF. We then fit a Normal or a SN to each of these CCFs, and the standard deviation of the distribution for the mean (RV or SN RV), the width parameter (FWHM or SN FWHM) and the asymmetric parameter (γ or BIS SPAN) is associated to the error on each of these parameters. [[**Referee:** It is not clear what you are doing here. I can understand bootstrapping a time series: you shuffle all the values keeping the time stamps fixed. How do you bootstrap a stellar spectrum? Do you shuffle the intensity values keeping the wavelengths fixed? That would produce a mess. Do you just add different levels of random noise? But the spectra already have noise in them, they are real observations. You need to state more clearly here what you did.]]

In the top plots of Fig. 13 we show the different errors for the RV, either defined as mean of the Normal (red triangles), mean of the SN (black circles) or median of the SN (cyan crosses), the width and the asymmetry of the CCFs for three star, HD215152, HD192310 and Corot-7, that are all at different SNR levels. The parameter SN50 corresponds to the SNR in order 50, which corresponds to a wavelength of 550 nm. In the bottom plots, we show the ratio between the parameters derived from the bootstrap analysis fitting the SN and the parameter derived from the bootstrap analysis fitting the Normal distribution. We first see that the errors on the CCF parameters only depends on the SNR and do not depend on the spectral type. This is true if the spectral type are not too different though, like here where we show the results for G and K dwarfs.

Concerning the standard errors related to the RVs, the ratio between the RV error measured by the bootstrap using the SN and Normal fitting is 1.6 when using the mean of the SN and 0.9 when using the median of the SN. In other words, by using the median of the SN as parameter for defining the radial velocity, we get standard errors 10% smaller than using the Normal fit.

Regarding the errors in width of the CCF, we see that the bootstrap analysis for the Normal or the SN are comparable. Therefore, the precision in CCF width is the same if we fit a Normal or a SN to the CCF.

Finally, for the errors in CCF asymmetry, we see that, when fitting the SN to the CCF, the asymmetry errors are 15% smaller. Therefore, the SN fit gives a better precision in CCF asymmetry than what can be reached using BIS SPAN. We recall moreover that, using the SN, all parameters are automatically retrieved in 1 single step, while in the common approach the RV and the BIS SPAN are calculated separately.

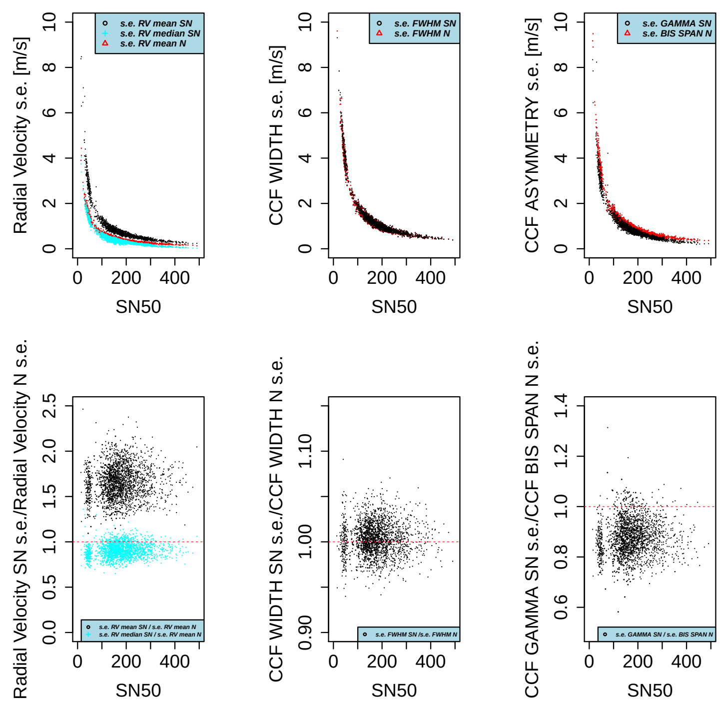

Figure 13: Comparison between the standard errors using the bootstrap analysis for the RVs, the FWHM and the asymmetry parameter. When using as parameter for the RV the mean of the SN (black circles), the standard errors are in average 60% larger than the standard errors retrieved fitting a Normal (red triangles). However, once the RV is defined as the median of the SN (cyan crosses), the standard errors are in average 10% smaller than the standard errors coming from the Normal fit. To use as asymmetry parameter the γ of the SN leads to standard errors in average 15% smaller than the standard errors related to the BIS SPAN. Note that for the asymmetry, the error in BIS SPAN is in $\mathrm{km\,s^{-1}}$. To be able to compare the errors in γ and BIS SPAN, we multiplied the error in γ by the slope of the correlation between γ and BIS SPAN.

6 Discussion

An analysis of the CCF residuals after fitting a Normal or SN distribution show that the SN is a slightly better model to explain the shape of the CCF. This comes from the fact that CCFs present a natural asymmetry due the convective blueshift.

[[**Umberto:** add discussions on SOAP example]]

We compared for five stars the difference between the RVs (defined as mean of a Normal, mean of the SN or median of the SN), FWHM and asymmetry (BIS SPAN in the Normal case and γ in the SN case). The γ parameter is linearly dependant on the BIS SPAN, with always a strong correlation coefficient. The slope of this linear correlation changes depending on the studied star. This is probably because the spectral type is different, therefore the effects from stellar activity are different.

When using as parameter for the RV the mean of the SN, the standard errors are in average 60% larger than the standard errors retrieved fitting a Normal. However, once the RV is defined as the median of the SN (cyan crosses), the standard errors are in average 10% smaller than the standard errors coming from the Normal fit. When looking at the correlation between the asymmetry and width parameters of the CCF (FWHM and BIS SPAN or the alternative indicators in Figueira et al. (2013) in the Normal case, and SN FWHM and γ in the SN case) with respect to the RVs (RVs in the Normal case or SN RVs in the SN case), we observe that the correlations are always stronger for the parameters of the SN. Therefore, the SN parameters are more sensitive to activity. In the case of Tau Ceti, which is at very low activity level, we find a significant correlation of 0.436 between γ and SN mean RV, while for all the other asymmetric parameterization, BIS SPAN or the alternative indicators in Figueira et al. (2013), the correlations are weaker with a maximum of 0.225.

7 Conclusion

In this paper we introduced a novel approach based on the Skew Normal (SN) distribution for deriving RVs and shape variations in the CCF of stars. When searching for small-mass exoplanets using the RV technique, it is essential to understand the shape variation of the CCF, which is a proxy for stellar activity effects. The standard approach consist at first to adjust a Normal distribution to the CCF to get the RV and FWHM, defined as the mean and the FWHM of the Normal distribution, and then to measure the asymmetry by calculating BIS SPAN. FWHM and BIS SPAN give us information on the line shape that are used to probe stellar activity signals.

In this paper we propose to conduct the analysis fitting a SN distribution to the CCF. Since the CCF presents a natural asymmetry due the convective blueshift, the SN distribution can better catch these aspects respect the Normal fit. On top of that, by using the SN distribution to fit CCFs, we can measure simultaneously the RV of the star, the width and the asymmetry of the CCF.

[[**Umberto:** add conclusions on SOAP example]]

Using the SN to fit CCFs brings a significant improvement in probing stellar activity. While for the Normal distribution mean and median are equivalent, using the SN fit different location

Star	# CCFs	R(SN γ, Bis-Span)	slope(SN γ, Bis-Span)	R(SN γ, SN mean RV)	R(Bis-Span, RV)	R(FIG BiGaussian, RV)	R(SN FWHM, SN mean RV)	R(FWHM, RV)
HD192310	1577	0.888	0.786	0.669(0.64; 0.695)	0.329(0.285; 0.373)	−0.333(−0.376; −0.289)	0.666(0.637; 0.692)	0.476(0.43670.514)
HD10700	7928	0.78	0.604	0.322(0.302; 0.342)	−0.073(−0.095; −0.0051)	0.127(0.105; 0.148)	0.421(0.403; 0.439)	0.529(0.513; 0.545)
HD215152	273	0.763	0.794	0.571(0.485; 0.646)	−0.067(−0.184; 0.052)	0.269(0.155; 0.376)	0.210(0.094; 0.321)	−0.138(−0.253; −0.020)
Corot 7	173	0.814	0.607	0.561(0.450; 0.656)	0.092(−0.058; 0.238)	−0.335(−0.228; −0.082)	−0.709(0.626; 0.776)	0.595(0.489; 0.683)

Table 2: Subset of notable correlations between the asymmetry parameter (and the FWHM) and the RVs for four stars: HD192310, HD10700, HD215152 and Corot 7.

parameters can be tested. While the median of the SN is more robust respect variations in the shape of the CCF, the mean of the SN is more sensible to changes in the asymmetry of the CCF. We suggest to use as parameter that defines the RV of the star the median of the SN, since the standard errors related to this parameter are 10% smaller than the standard errors retrieved using the Normal distribution. For evaluating changes in the asymmetry of the CCF, we suggest to use the mean of the SN. The correlations between SN mean RV and SN FWHM, and SN mean RV and γ (the asymmetry parameter of the SN) are much stronger than the correlations between the equivalent parameters derived using a Normal fit (RV, FWHM and BIS SPAN or the asymmetric parameters described in Figueira et al. (2013)). The precision on the asymmetry measured by γ is greater than the one on BIS SPAN by ∼15%. Therefore when searching for rotational periods in the data, or applying Gaussian Processes to account for stellar activity signals, the SN parameters should be used.

Finally, we also encourage the use of bootstrapping to estimate more realistic errors on the different parameters of the Normal or SN fitted to the CCF, mainly in the low SNR regime where a gain of 50% can be reached. This takes significantly more time, but note that 100 realization are enough to get a good estimation of errors.

8 Acknowledgements

We are grateful to all technical and scientific collaborators of the HARPS Consortium, ESO Headquarters and ESO La Silla who have contributed with their extraordinary passion and valuable work to the success of the HARPS project. XD is grateful to the Society in Science–The Branco Weiss Fellowship for its financial support.

A Appendix

In this Appendix we summarized the analyses run on other 4 stars: HD192310, HD10700, HD215152 and finally Corot-7. Table 2 outlines the comparison between the results obtained by the SN fits and the ones based on the Normal distribution. The results are all consistent with the conclusions presented in the paper. The correlation between γ and the parameter SN mean RV is stronger than the correlation between the BIS SPAN and the mean of the Normal for all the considered stars.

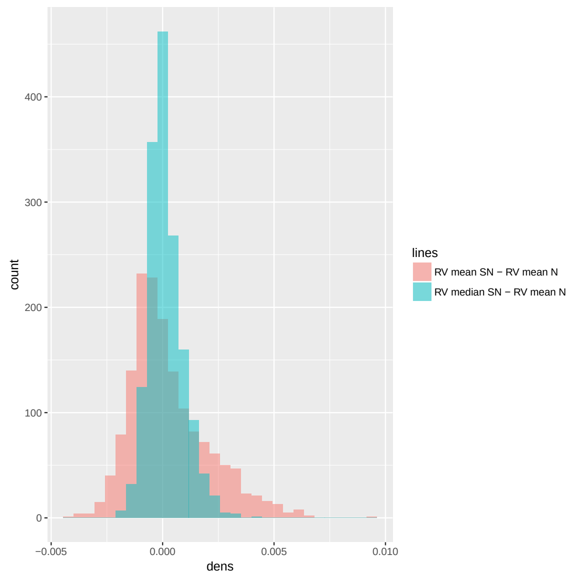

Figure 14: RVs comparison for HD192310 considering a Normal and a SN fitting (using both the mean and the median).

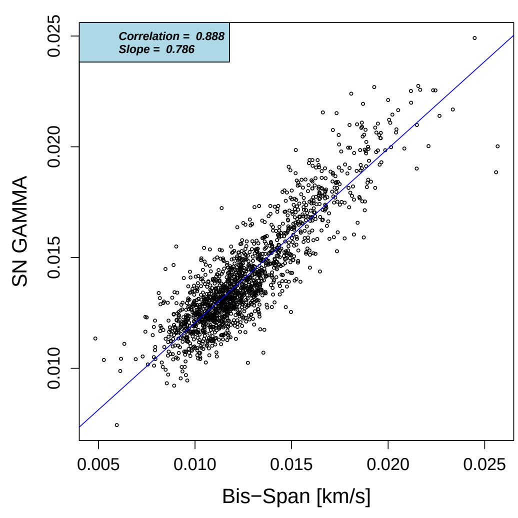

Figure 15: Correlation between γ and the BIS SPAN for HD192310.

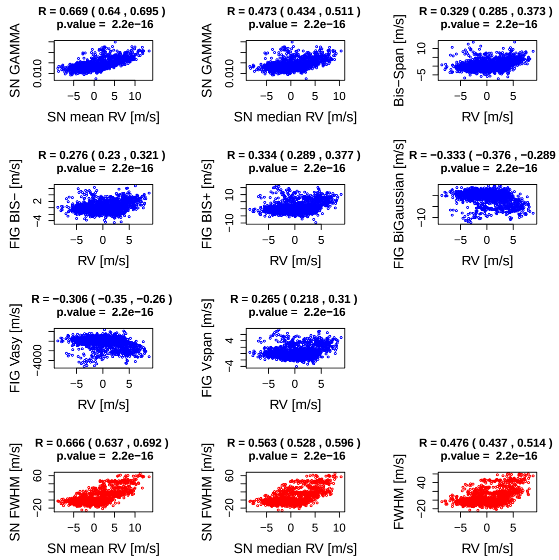

Figure 16: Correlation between the asymmetry parameters and the RVs for HD 192310. The last two plots show the correlation between the FWHM and the RVs for HD 192310 using respectively the SN and the Normal fits.

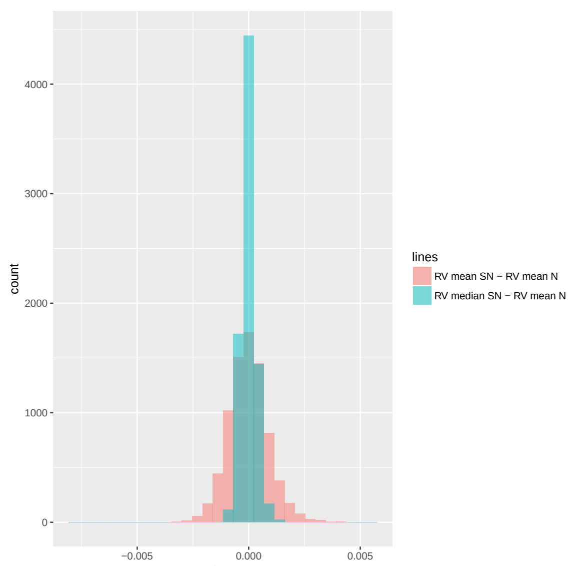

Figure 17: RVs comparison for Tau Ceti considering a Normal and a SN fitting (using both the mean and the median).

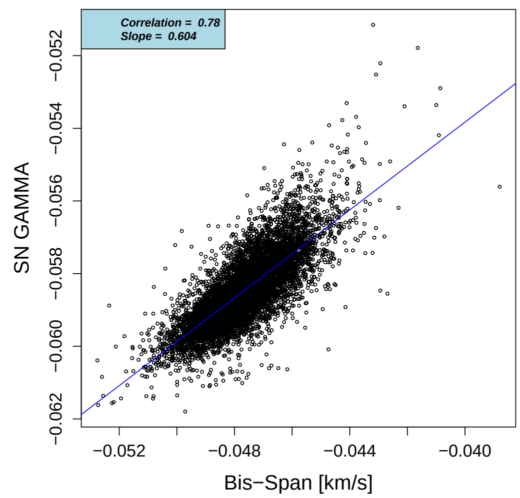

Figure 18: Correlation between γ and the BIS SPAN for Tau Ceti.

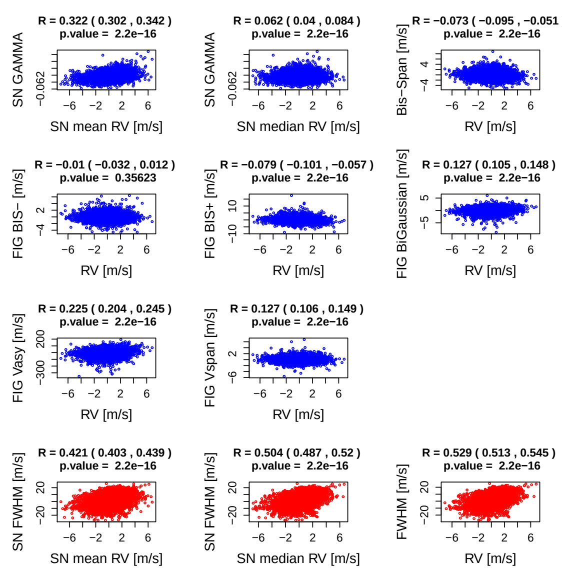

Figure 19: Correlation between the asymmetry parameters and the RVs for Tau Ceti. The last two plots show the correlation between the FWHM and the RVs for Tau Ceti using respectively the SN and the Normal fits.

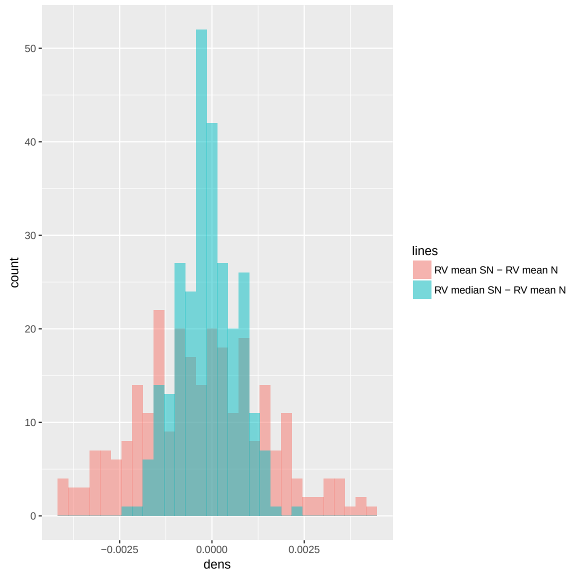

Figure 20: RVs comparison for HD215152 considering a Normal and a SN fitting (using both the mean and the median).

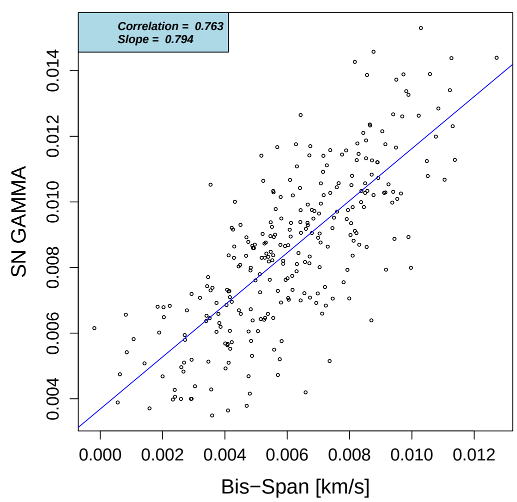

Figure 21: Correlation between γ and the BIS SPAN for HD 215152.

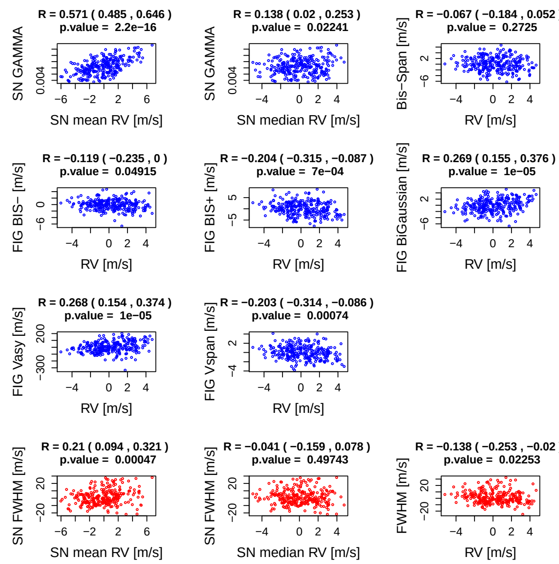

Figure 22: Correlation between the asymmetry parameters and the RVs for HD 215152. The last two plots show the correlation between the FWHM and the RVs for HD 215152 using respectively the SN and the Normal fits.

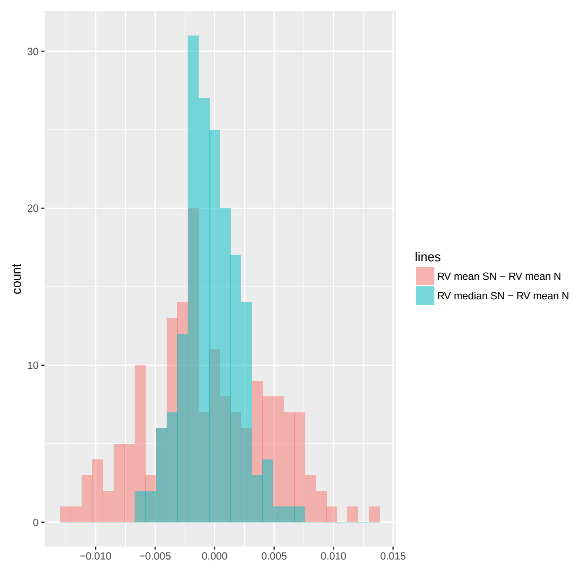

Figure 23: RVs comparison for Corot-7 considering a Normal and a SN fitting (using both the mean and the median).

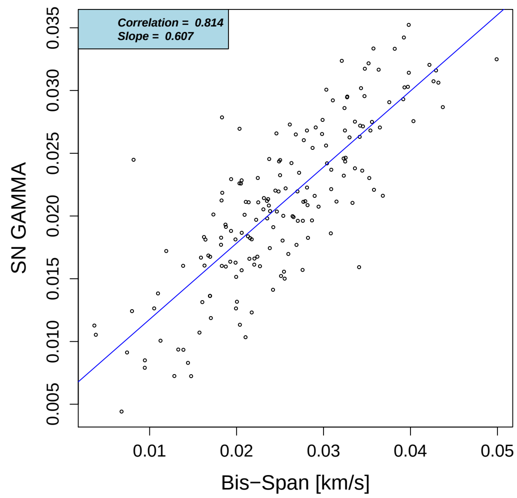

Figure 24: Correlation between γ and the BIS SPAN for Corot-7.

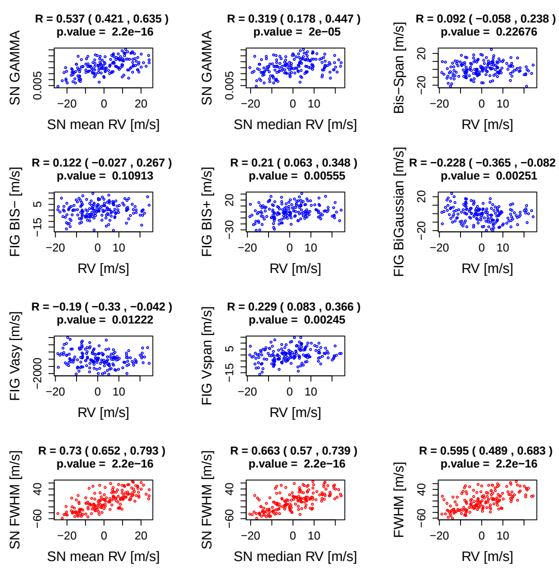

Figure 25: Correlation between the asymmetry parameters and the RVs for Corot-7. The last two plots show the correlation between the FWHM and the RVs for Corot-7 using respectively the SN and the Normal fits.

References

G. Anglada-Escudé and R. P. Butler. The harps-terra project. i. description of the algorithms, performance, and new measurements on a few remarkable stars observed by harps. *The Astrophysical Journal Supplement Series*, 200(2):15, 2012.

A. Azzalini. A class of distributions which includes the normal ones. *Scandinavian journal of statistics*, pages 171–178, 1985.

A. Azzalini and A. Capitanio. The skew-normal and related families. institute of mathematical statistics monographs, 2014.

A. Baranne, D. Queloz, M. Mayor, G. Adrianzyk, G. Knispel, D. Kohler, D. Lacroix, J.-P. Meunier, G. Rimbaud, and A. Vin. Elodie: A spectrograph for accurate radial velocity measurements. *Astronomy and Astrophysics Supplement Series*, 119(2):373–390, 1996.

I. Boisse, X. Bonfils, and N. Santos. Soap-a tool for the fast computation of photometry and radial velocity induced by stellar spots. *Astronomy & Astrophysics*, 545:A109, 2012.

F. Bouchy, R. Díaz, G. Hébrard, L. Arnold, I. Boisse, X. Delfosse, S. Perruchot, and A. Santerne. Sophie+: First results of an octagonal-section fiber for high-precision radial velocity measurements. *Astronomy & Astrophysics*, 549:A49, 2013.

F. Cavallini, G. Ceppatelli, and A. Righini. Asymmetry and shift of three fe i photospheric lines in solar active regions. *Astronomy and Astrophysics*, 143:116–121, 1985.

R. Cosentino, C. Lovis, F. Pepe, A. C. Cameron, D. W. Latham, E. Molinari, S. Udry, N. Bezawada, M. Black, A. Born, et al. Harps-n: the new planet hunter at tng. In *Proc. SPIE*, volume 8446, page 84461V, 2012.

M. Desort, A.-M. Lagrange, F. Galland, S. Udry, and M. Mayor. Search for exoplanets with the radial-velocity technique: quantitative diagnostics of stellar activity. *Astronomy & Astrophysics*, 473(3):983–993, 2007.

D. Dravins, L. Lindegren, and Å. Nordlund. Solar granulation-influence of convection on spectral line asymmetries and wavelength shifts. *Astronomy and Astrophysics*, 96:345–364, 1981.

X. Dumusque, F. Pepe, C. Lovis, D. Ségransan, J. Sahlmann, W. Benz, F. Bouchy, M. Mayor, D. Queloz, N. Santos, et al. An earth-mass planet orbiting [agr] centauri b. *Nature*, 491(7423):207–211, 2012.

X. Dumusque, I. Boisse, and N. Santos. Soap 2.0: A tool to estimate the photometric and radial velocity variations induced by stellar spots and plages. *The Astrophysical Journal*, 796(2):132, 2014.

P. Figueira, N. Santos, F. Pepe, C. Lovis, and N. Nardetto. Line-profile variations in radial-velocity measurements-two alternative indicators for planetary searches. *Astronomy & Astrophysics*, 557:A93, 2013.

A. M. Fiorenzano, R. Gratton, S. Desidera, R. Cosentino, and M. Endl. Line bisectors and radial velocity jitter from sarg spectra. *Astronomy & Astrophysics*, 442(2):775–784, 2005.

D. F. Gray. The third signature of stellar granulation. *The Astrophysical Journal*, 697(2):1032, 2009.

A. P. Hatzes. Simulations of stellar radial velocity and spectral line bisector variations: I. nonradial pulsations. *Publications of the Astronomical Society of the Pacific*, 108(728):839, 1996.

A. P. Hatzes. Starspots and exoplanets. *Astronomische Nachrichten*, 323(3-4):392–394, 2002.

M. Kurster, M. Endl, F. Rouesnel, S. Els, A. Kaufer, S. Brillant, A. Hatzes, S. Saar, and W. Cochran. The low-level radial velocity variability in barnard's star (= gj 699). secular acceleration, indications for convective redshift, and planet mass limits. *ASTRONOMY AND ASTROPHYSICS-BERLIN-*, 403(3): 1077–1088, 2003.

A.-M. Lagrange, M. Desort, and N. Meunier. Using the sun to estimate earth-like planets detection capabilities-i. impact of cold spots. *Astronomy & Astrophysics*, 512:A38, 2010.

L. Lindegren and D. Dravins. The fundamental definition of ?radial velocity? *Astronomy & Astrophysics*, 401(3):1185–1201, 2003.

N. Meunier, M. Desort, and A.-M. Lagrange. Using the sun to estimate earth-like planets detection capabilities-ii. impact of plages. *Astronomy & Astrophysics*, 512:A39, 2010.

F. Pepe, M. Mayor, F. Galland, D. Naef, D. Queloz, N. Santos, S. Udry, and M. Burnet. The coralie survey for southern extra-solar planets vii-two short-period saturnian companions to hd 108147 and hd 168746. *Astronomy & Astrophysics*, 388(2):632–638, 2002.

C. PHASE. Setting new standards with harps. *The Messenger*, 114:20, 2003.

D. Queloz, M. Mayor, L. Weber, A. Blécha, M. Burnet, B. Confino, D. Naef, F. Pepe, N. Santos, and S. Udry. The coralie survey for southern extra-solar planets. i. a planet orbiting the star gliese 86. *Astronomy and Astrophysics*, 354:99–102, 2000.

D. Queloz, G. Henry, J. Sivan, S. Baliunas, J. Beuzit, R. Donahue, M. Mayor, D. Naef, C. Perrier, and S. Udry. No planet for hd 166435. *Astronomy & Astrophysics*, 379(1):279–287, 2001.

A. Quirrenbach, P. Amado, J. Caballero, R. Mundt, A. Reiners, I. Ribas, W. Seifert, M. Abril, J. Aceituno, F. Alonso-Floriano, et al. Carmenes instrument overview. In *Proc. SPIE*, volume 9147, page 91471F, 2014.

S. H. Saar and R. A. Donahue. Activity-related radial velocity variation in cool stars. *The Astrophysical Journal*, 485(1):319, 1997.

A. Thompson, C. Watson, E. de Mooij, and D. Jess. The changing face of α centauri b: probing plage and stellar activity in k dwarfs. *Monthly Notices of the Royal Astronomical Society: Letters*, 468(1):L16–L20, 2017.

R. G. Tull. High-resolution fiber-coupled spectrograph of the hobby-eberly telescope. *Proc. Soc. Photo-opt. Inst. Eng.*, 3355:387, 1998.

R. G. Tull, P. J. MacQueen, C. Sneden, and D. L. Lambert. The high-resolution cross-dispersed echelle white pupil spectrometer of the mcdonald observatory 2.7-m telescope. *Publications of the Astronomical Society of the Pacific*, 107(709):251, 1995.

S. Vogt. *PASP*, (99):1214, 1987.

S. S. Vogt, S. L. Allen, and B. C. Bigelow. *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, 2198(362), 1994.

H.-H. Voigt. " drei-strom-modell" der sonnenphotosphäre und asymmetrie der linien des infraroten sauerstoff-tripletts. mit 12 textabbildungen. *Zeitschrift fur Astrophysik*, 40:157, 1956.